Introduction à la théorie des codes correcteurs d'erreurs

Cours 1 : Généraillés

Odile PAPINI

POLYTECH
Université d'Aix-Marseille
odile.papini@univ-amu.fr
http://odile.papini.perso.esil.univmed.fr/sources/CODAGE.html

Plan du cours

Bibliographie I


F. J. Mac Willians & N. J. A. Sloane
The Theory of Error Correcting codes.
North Holland Publising, ed. 1978.


O. Papini & J. Wolfmann
Algèbre discrète et codes correcteurs d'erreurs.
Springer Verlag ed. 1995.


Support de cours
Claude Carlet Université Paris 13
http://www.math.univ-paris13.fr/ schartz/Mali/Mali07/ccc.pdf

Problématique

Transmission d'information

évolution : analogique → numérique

apparition d'erreurs de transmission :

- télécommunication à grande distance
- perturbations météorologiques
- dépôt de poussière

Objet de la théorie des codes correcteurs d'erreurs

Protéger les transmissions des erreurs, de façon efficace

Notion d'information

théorie des codes correcteurs d'erreurs $\rightarrow$ théorie de l'information

notion moderne d'information

distinction entre :

- **sens d'un message**
- **forme d'un message**

émergence d'une **théorie de l'information** C. Shannon (1948)

traitement automatique de l'information sur lequel s'appuie la plupart des moyens modernes de communication

Définition mathématique de l'information

- support de l'information numérique : système à 2 états d'équilibre 0 et 1
- information élémentaires : **bits** (**bi**nary digi**t**)
- toute information est représentée par une combinaison d'informations élémentaires
- quantité d'information : logarithme du nombre de choix possibles
- unité de quantité d'information : **shannon (sh)** (normes CCITT)

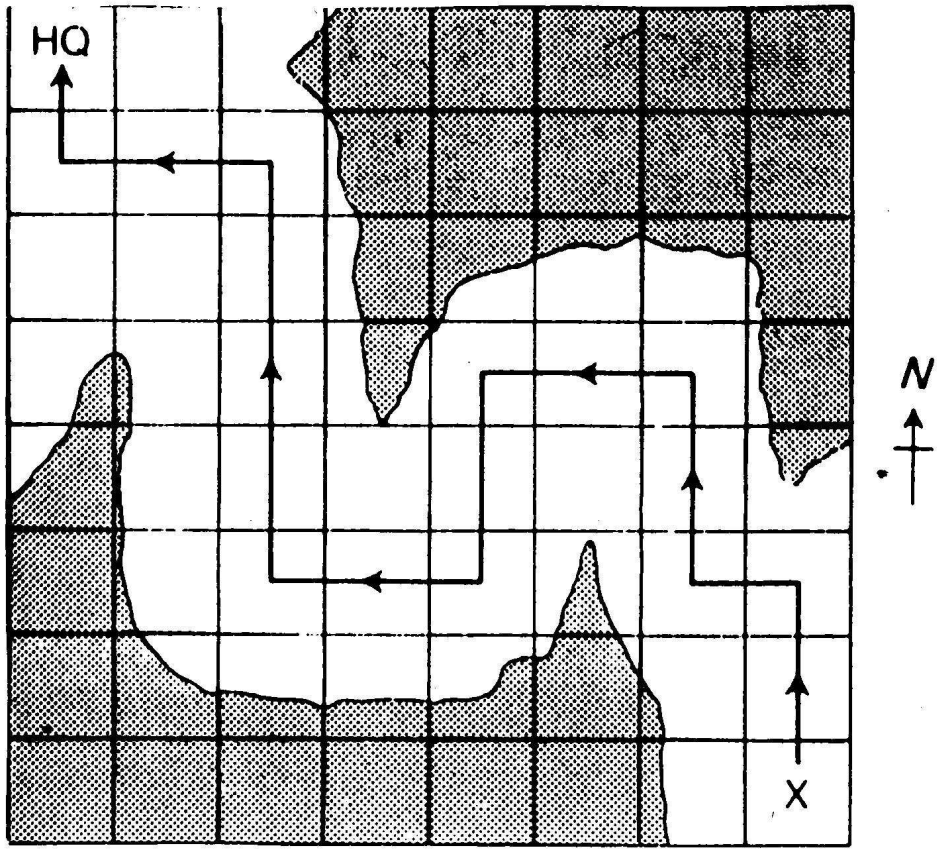
  quantité d'information apportée par la connaissance d'un état parmi 2 états possibles : $log_2(2)$

information I : combinaison de m bits

quantité d'information I : $log_2(2^m)$

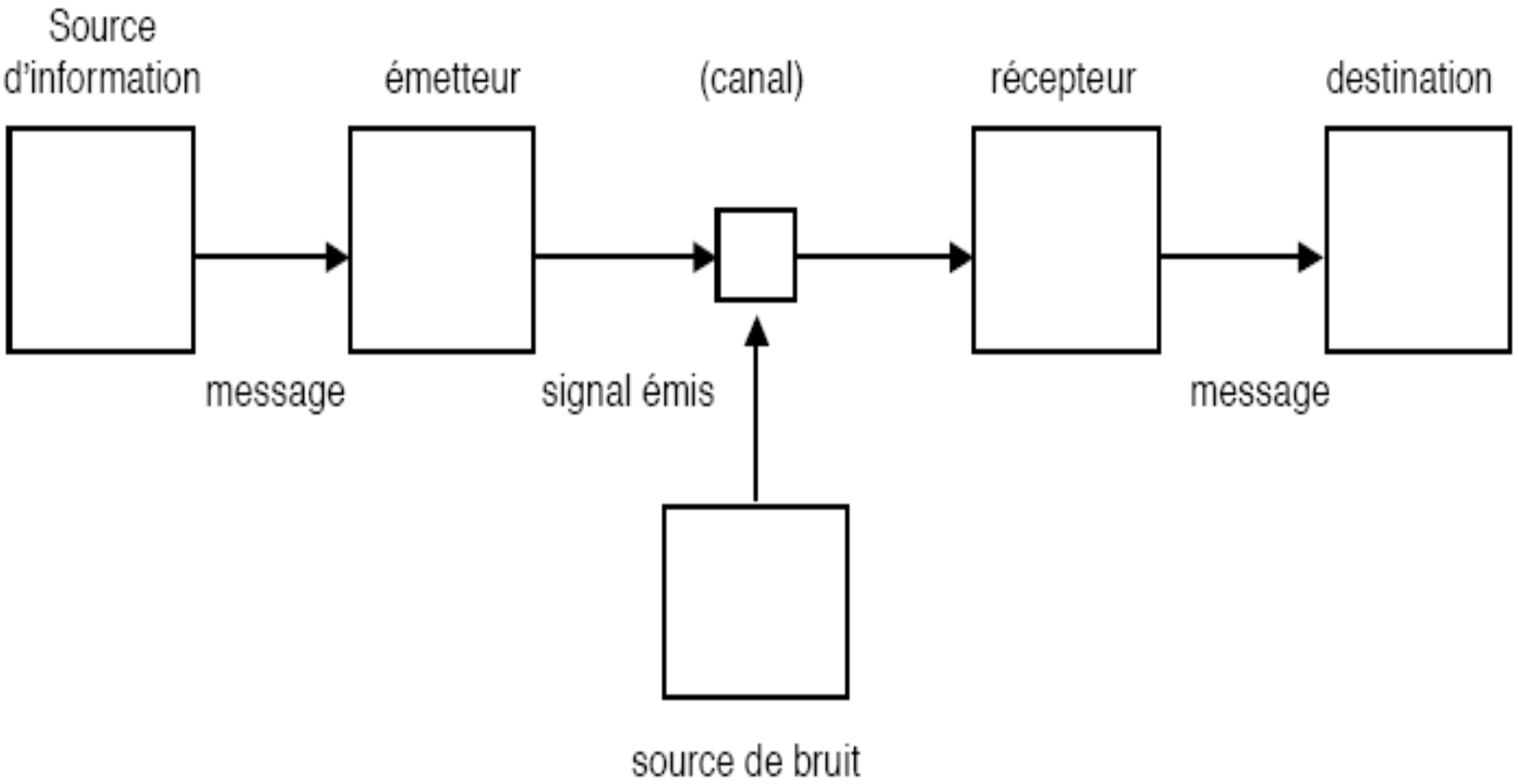
Un exemple introductif

Codage d'un itinéraire selon les directions cardinales



Théorie de l'information

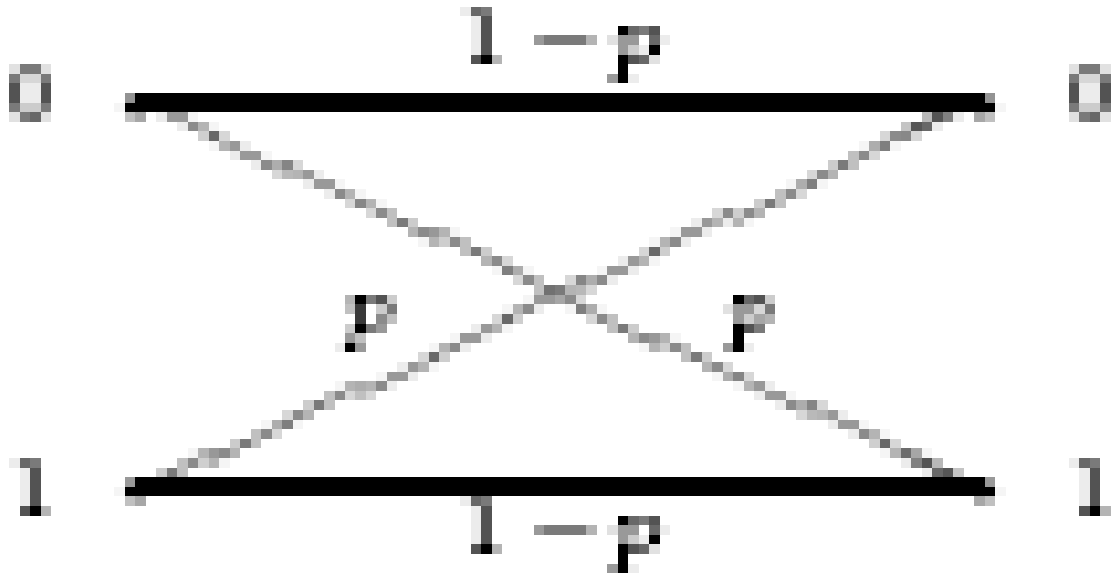
Description et étude des systèmes de communication : Claude Shannon 1948



Canal de transmission

canal **binaire symétrique**

p : probabilité qu'un 1 soit reçu lorsqu'un 0 est transmis
ou probabilité qu'un 0 soit reçu lorsqu'un 1 est transmis



Canal de transmission

canal **binaire symétrique**

p : probabilité qu'un 1 soit reçu lorsqu'un 0 est transmis
ou probabilité qu'un 0 soit reçu lorsqu'un 1 est transmis

$$0 \xrightarrow{p} 1 \quad \Big| \quad 0 \xrightarrow{1-p} 0 \qquad\qquad 1 \xrightarrow{p} 0 \quad \Big| \quad 1 \xrightarrow{1-p} 1$$

quantité d'information transmise par le canal : **capacité**

$$C(p) = 1 - (p\ log_2(\frac{1}{p}) + (1-p)\ log_2(\frac{1}{1-p}))$$

$$C(p) = 1 + p\ log_2(p) + (1-p)\ log_2(1-p)$$

Théorème de Shannon

théorème

Soient

$C(p)$ la capacité d'un canal de transmission

2 valeurs arbitraires $\delta > 0$ et $R < C(p)$

k le nombre de symboles du message à transmettre

P la probabilité de décoder incorrectement un mot reçu

Il existe un code de longueur n suffisamment grande tel que

$$\frac{k}{n} \geq R \quad \text{et} \quad P < \delta$$

Problématique

2-ième théorème de Shannon :

Le codage est une opération efficace au prix d'une certaine redondance

- Trouver de "bons codes"
- Trouver des méthodes de décodage efficaces

technologie mécanique

communication à distance

préoccupation tardive dans l'histoire de l'humanité

fin du *XVIII*ième siècle :

réseau de communication à distance systématique

- technologie mécanique
- 1794 : premier **télégraphe aérien** de Claude Chappe

technologie électrique

- 1832 : **télégraphe électrique** de Samuel F.B Morse
- 1866 : premier **câble transatlantique**
- 1874 : Thomas Edison perfectionne le télégraphe
- télégraphe de E. Baudot avec un code binaire,
  - **le code à cinq moments**
    $\rightarrow$ code RTTY pour les transmissions par telex)

technologie électromagnétique

- 1875 : **téléphone** Alexandre Graham Bell
- premier système de **transmission sans fil**, TSF, par A. Popov
- 1896 : première installation, G. Marconi : début de la **radiocommunication**

technologie électronique

mise en évidence l'émission thermo-ionique (T. Edison)

passage des électrons d'une électrode négative à une électrode positive

- 1902 : **diode** A. Fleming
- 1906 : **triode** L. De Forest

conception de tubes électroniques : début de l'ère électronique

- 1917 : premiers postes émetteurs-récepteurs portatifs à lampes

matrise des procédés de modulation des ondes

naissance de la radiodiffusion

- 1920 : premiers essais aux Etats Unis
- 1921 : émissions régulières en France à partir de la tour Eiffel
- 1930 : développement de la modulation de fréquence (bande FM)

technologie électronique

L'effet photo-électrique

éjection d'électrons hors des atomes au moyen de photons de lumière

- 1905 : cellule **photoélectrique** Elster et Geitel.
- **tube cathodique** rend possible la transmission d'images
- 1923 : premiers essais de **télévision** en Angleterre
- 1938 : diffusion d'émissions en France à partir de la tour Eiffel

recherches sur les matériaux semi-conducteurs

- 1947 : premier **transistor** W. Shockley, W. H. Brattain et J. Barden (laboratoires de la Bell Telephone)

télécommunications et réseaux

évolution des télécommunications

- contexte de guerre froide : course à l'espace
- 1960 : satellites comme relais pour les télécommunications
- Echo 1960, Telstar 1962
- aujourd'hui : satellites géostationnaires

évolution et le développement de l'informatique

- réseaux d'ordinateurs
- réseau de télécommunication : RNIS
- 1987 : NUMERIS de France-Télécom.

évolution de la transmission de l'information

analogique $\rightarrow$ numérique

codage de l'information

- **code binaire** : Francis Bacon (1561 – 1628) système de codage : 2 opérations
  - codage de chaque lettre de l'alphabet par une combinaison de 2 symboles
  - pour chaque symbole une typographie différente.

- G. W. von Liebniz (1643 – 1716) : notation binaire pour les nombres.

- J. M. Jacquard (1752 – 1834) : codage binaire (cartes perforées)

Exemple du codage de F. Bacon (cryptographie)

- codage de la lettre a : *aaaaa*
- codage de la lettre b : *aaaba*

  ...,
- codage de la lettre z : *babbb*

symbole *a* : typographie minuscule symbole *b* : une typographie majuscule
message à transmettre :

non

codé par :

[*abbaa*][*abbab*][*abbaa*]

vOUs poUVeZ vENir

code Morse

Morse 1838 : codage de chaque lettre de l'alphabet par une combinaison de 3 symboles (code ternaire)

Lettres	Morse	Lettres	Morse
A	.–	*N*	–.
B	–...	*O*	– – –
C	–. – .	*P*	. – –.
D	–..	*Q*	– – .–
E	.	*R*	. – .
F	.. – .	*S*	...
G	– – .	*T*	–
H	....	*U*	..–
I	..	*V*	...–
J	. – ––	*W*	. – –
K	–.–	*X*	–..–
L	. – ..	*Y*	–. – –
M	––	*Z*	– – ..

codage de l'information

pour les codes présentés :

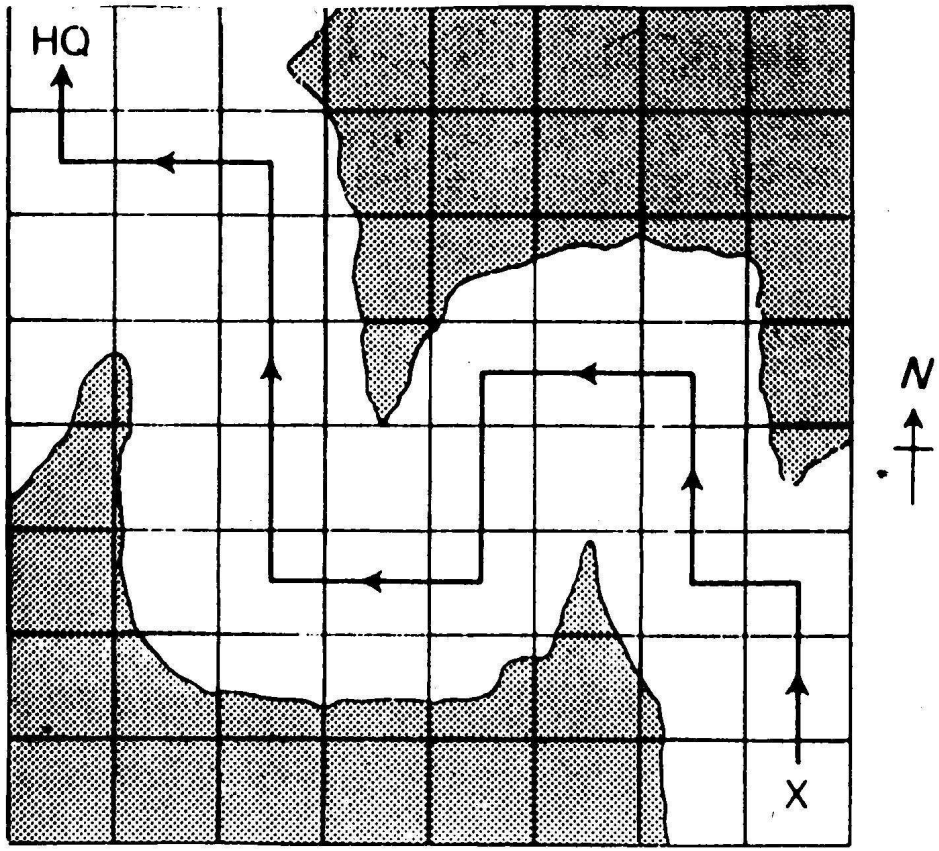
- **représentation** de l'information sous forme symbolique
- **pas de protection** contre une éventuelle perte d'information
- **n'ecessité d'introduction de symboles supplémentaires**, ou **redondants**,

exemple : langage des aviateurs

C : C**HARLIE**
P : P**APA**
T : T**ANGO**

Un exemple introductif

Codage d'un itinéraire selon les directions cardinales



Préhistoire des codes correcteurs d'erreurs

- **laboratoires de la Bell Telephone** 1937 – 1939 : premières recherches sur l'utilisation des codes détecteurs et correcteurs d'erreurs

- **seconde guerre mondiale :** recherches activement développées
  - but : concevoir un calculateur adapté aux problèmes de commandes de tirs pour le "National Defense Research Council"
  - "l'interpolateur à relais" 1943 : utilisation de codes détecteurs d'erreurs

- 1947 W. Hamming : **premier code corrigeant une seule erreur**

Code

A : **Alphabet**

$(x_1, \cdots, x_n)$: **un n-uplet** avec $x_i \in A$

mot : un n-uplet $(x_1, \cdots, x_n)$

longueur du mot : l'entier n

code

l'ensemble des mots fabriqués sur A

L'alphabet le plus utilisé : $\{0, 1\}$

Problématique

A partir d'une telle représentation :

- **décrire les propriétés de transmission** en vue de **détecter et de corriger des erreurs** apparues au cours de la transmission des messages
- **inversement choisir convenablement l'alphabet et le code** pour tenter d'améliorer la transmission
  - en particulier, choisir un alphabet et un code dotés de **structures algébriques et combinatoires** particulières

Distance de Hamming

A : ensemble fini, non vide et n un entier naturel
A^n : l'ensemble des $x = (x_1, \cdots, x_n)$ avec $x_i \in A$

distance de Hamming entre x et y :

nombre de composantes pour lesquelles x et y diffèrent.

si $x = (x_1, x_2, ...x_n)$ et $y = (y_1, y_2, ...y_n)$, alors :

$$d(x, y) = \sharp\{i \in \{1, 2, \ldots, n\},\ \text{tq } x_i \neq y_i\}$$

exemple

$A = \{+, ?, \neq\}$, $n = 5$, $\quad x = (?, ?, +, ?, \neq)$ et $y = (?, \neq, \neq, +, \neq)$

$d(x, y)$?

Distance de Hamming

La distance de Hamming est bien une distance

$\forall x, y, z \in A^n$ on a :

i) $d(x, y) \in \mathbb{R}^+$;

ii) $d(x, y) = 0 \iff x = y$;

iii) $d(x, y) = d(y, x)$;

iv) $d(x, y) \leq d(x, z) + d(z, y)$.

(démontrastion laissée en exercice).

Code

définition d'un code

Soit A un alphabet.

Un *code* sur A de longueur n est un sous-ensemble C de A^n.

les éléments de C sont appelés les *mots* du code.

exemple

$$C = \{(0,1,1,0,1,0),(1,1,1,0,1,1),(1,1,1,1,1,1),\\(1,1,0,0,0,0),(0,1,1,1,0,0)\}$$

est un code de longueur 6 sur l'alphabet $A = \{0,1\}$.

Codes correcteurs

Lorsqu'on utilise un code pour transmettre des messages
$x = (x_1, x_2, ...x_n)$ est un **mot envoyé**
$x' = (x'_1, x'_2, ...x'_n)$ le **mot reçu** éventuellement **entaché d'erreurs**

$$d(x, x') : \textbf{nombre d'erreurs commises}$$

Lorsqu'**on suppose** qu'**il n'y a pas plus de** e **erreurs commises**

$$d(x, x') \leq e$$

on pourra **corriger** ces erreurs

corriger ces erreurs :

retrouver x**, à partir de** x' à condition que **chaque mot erroné reçu ne puisse provenir que d'un seul mot du code.**

Condition de décodage d'ordre e

condition de décodage d'ordre e

Un code C de longueur n sur un alphabet A vérifie la condition de décodage d'ordre e si **pour tout** x' de A^n, **il existe au plus un mot** x de C tel que

$$d(x, x') \leq e$$

Cette condition est équivalente à :

les boules fermées de rayon e, centrées sur les mots de C, soient deux à deux disjointes.

Exemples

Ces codes vérifient-t-ils la condition de décodage d'ordre e ?

exemple 1 : $A = \{0, 1\}$, $n = 5$, $e = 1$,

$$C = \{x = (0, 1, 1, 1, 0), y = (1, 0, 1, 0, 1), z = (1, 1, 0, 1, 1)\}$$

exemple 2 : $A = \{0, 1\}$, $n = 3$, $e = 1$,

$$C = \{x = (0, 0, 0), y = (1, 0, 1)\}$$

Distance minimale d'un code

définition

La **distance minimale** d'un code C est **la plus petite des distances** non nulles entre les mots du code

$$d_{min} = inf\{d(x,y),\ (x,y) \in C \times C,\ x \neq y\}$$

théorème

Soit d la distance minimale d'un code C

si $d \geq 2e+1$ **alors** C **vérifie la condition de décodage d'ordre** e

Distance minimale d'un code

- si d est pair, $d = 2r$ alors le meilleur e pour vérifier la condition de décodage est $e = r - 1$
- si d est impair, $d = 2r + 1$ alors le meilleur e pour vérifier la condition de décodage est $e = r$

définition capacité de correction : $e \in \mathbb{N}$

un code de distance minimale d est dit e-correcteur si

$$\lceil \frac{d-1}{2} \rceil = e$$

Code parfait

Un code C de longueur n sur un alphabet A

définition :rayon de recouvrement

Le rayon de recouvrement $\rho(C)$ de C estle **plus petit rayon** r tel que l'ensemble des boules de rayon r centrées sur chaque mot du code C forme **un recouvrement** de A^n

définition: code parfait

Un code C est **parfait** si $e = \rho(C)$

Questions

Un code C de longueur n sur un alphabet A

- Quel est le nombre **maximum de mots** que peut contenir un code C de **capacité de correction** e ?
- Quel est le nombre **minimum de mots** que doit contenir un code C pour avoir **un rayon de recouvrement** $\rho(C)$ fixé ?

Réponses: Borne d'empilement de sphères

Proposition 1

Soit C un code de longueur n sur un alphabet A, pour tout entier $r \leq n$ on a:

$$\forall x \in A^n, \quad |B(x,r)| = \Sigma_{i=0}^{r} (|A|-1)^i \ C_n^i$$

Proposition 2 : borne d'empilement de sphères

Soit C un code de longueur n sur un alphabet A, de capacité de correction e on a:

$$|C| \ \Sigma_{r=0}^{e} (|A|-1)^r \ C_n^r \ \leq \ A^n$$

Codes équivalents

codes équivalents

deux codes sont **équivalents** si l'un d'eux est obtenu à partir de l'autre en appliquant à chaque mot une même permutation des composantes

définition

Soit C et C' des codes de longueur n sur un alphabet A
soit σ une application de $\{1, 2, \cdots, n\}$ et soit l'application $\bar{\sigma}$:

$$\begin{array}{ccc} A^n & \rightarrow & A^n \\ (x_1, \cdots, x_n) & \rightarrow & (x_{\sigma(1)}, \cdots, x_{\sigma(n)}) \end{array}$$

C et C' sont **équivalents** si il existe σ telle que $\bar{\sigma}(C)$